Thermodynamics Of Neutral Protein Evolution

Jesse D. Bloom,[*] Alpan Raval,[†] Claus O. Wilke[‡]

February 5, 2008

[*]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125

[†]Keck Graduate Institute of Applied Life Sciences and School of Mathematical Sciences, Claremont Graduate University, Claremont, CA 91711

[‡]Section of Integrative Biology and Center for Computational Biology and Bioinformatics, University of Texas at Austin, Austin, TX 78712

Running Head: Thermodynamics Of Evolving Proteins

Key Words: protein stability, mutational robustness, index of dispersion, molecular clock, lattice protein

Corresponding Author:

Jesse D. Bloom

Division of Chemistry and Chemical Engineering

California Institute of Technology

Mail Code 210-41

Pasadena, CA 91125

(626) 354-2565 (ph.)

(626) 568-8743 (fax)

jesse.bloom@gmail.com

Abstract

Naturally evolving proteins gradually accumulate mutations while continuing to fold to stable structures. This process of neutral evolution is an important mode of genetic change, and forms the basis for the molecular clock. We present a mathematical theory that predicts the number of accumulated mutations, the index of dispersion, and the distribution of stabilities in an evolving protein population from knowledge of the stability effects ($\Delta\Delta G$ values) for single mutations. Our theory quantitatively describes how neutral evolution leads to marginally stable proteins, and provides formulae for calculating how fluctuations in stability can overdisperse the molecular clock. It also shows that the structural influences on the rate of sequence evolution observed in earlier simulations can be calculated using just the single-mutation $\Delta\Delta G$ values. We consider both the case when the product of the population size and mutation rate is small and the case when this product is large, and show that in the latter case the proteins evolve excess mutational robustness that is manifested by extra stability and an increase in the rate of sequence evolution. All our theoretical predictions are confirmed by simulations with lattice proteins. Our work provides a mathematical foundation for understanding how protein biophysics shapes the process of evolution.

INTRODUCTION

Proteins evolve largely through the slow accumulation of amino acid substitutions. Over evolutionary time, this process of sequence divergence creates homologous proteins that differ at the majority of their residues, yet still fold to similar structures that often perform conserved biochemical functions (LESK and CHOTHIA 1980). The maintenance of structure and function during sequence divergence suggests that much of protein evolution is neutral in the sense that observed sequence changes frequently do not alter a protein's ability to fold and adequately perform the biochemical function necessary to enable its host organism to survive. This comparative evidence for neutrality in protein evolution has been corroborated by experimental studies showing that the mutations separating diverged sequences often have no effect other than modest and additive changes to stability (SERRANO *et al.* 1993), and that a large fraction of random mutations do not detectably alter a protein's structure or function (SHORTLE and LIN 1985; PAKULA *et al.* 1986; LOEB *et al.* 1989; GUO *et al.* 2004; BLOOM *et al.* 2005; BLOOM *et al.* 2006). In this respect, it seems that protein evolution should be well described by Kimura's neutral theory of evolution, which holds that most genetic change is due to the stochastic fixation of neutral mutations (KIMURA 1983). One of the key predictions of the neutral theory is that assuming a constant mutation rate, the number of mutations separating two proteins should be proportional to the time since their divergence (KIMURA 1983). Indeed, the observation by ZUCKERKANDL and PAULING (1965) that proteins are "molecular clocks" that accumulate mutations at a roughly constant rate has long been taken as one of the strongest pieces of evidence supporting the neutral the-

ory (OHTA and KIMURA 1971).

However, mutations that are neutral with respect to a protein's capacity to perform its biological function often affect protein thermodynamics. The biological functions of most proteins depend on their ability to fold to thermodynamically stable native structures (ANFINSEN 1973). Yet natural proteins are typically only marginally stable, with free energies of folding (ΔG_f) between -5 and -15 kcal/mol (FERSHT 1999). Most random mutations to proteins are destabilizing (GODOY-RUIZ *et al.* 2004; PAKULA *et al.* 1986; MATTHEWS 1993; KUMAR *et al.* 2006), and their effects on stability (measured as $\Delta\Delta G$, the ΔG_f of the mutant protein minus the ΔG_f of the wildtype protein) are frequently of the same magnitude as a protein's net stability. The impact of a mutation on a protein's function can therefore depend on the protein's stability: a moderately destabilizing mutation that is easily tolerated by a stable parent protein may completely disrupt the folding of a less stable parent. This effect of protein stability on mutational tolerance has been verified by experiments demonstrating that more stable protein variants are markedly more robust to random mutations (BLOOM *et al.* 2005; BLOOM *et al.* 2006).

The fact that mutations that are neutral with respect to direct selection for protein function can affect a protein's tolerance to subsequent mutations is not consistent with the simplest formulation of the neutral theory of evolution, which tends to assume that the fraction of mutations that is neutral remains constant in time. KIMURA (1987) himself recognized the possibility that the neutrality might change, and TAKAHATA (1987) mathematically treated the consequences of a "fluctuating neutral space". In particular, Takahata showed that fluctuating neutrality could explain the observed overdispersion in the molecular clock (CUTLER 2000b) (the tendency for

5

the variance in the number of fixed mutations to exceed the expectation for the Poisson process predicted by the neutral theory) long considered troublesome for the neutral theory. However, further progress on this topic was stymied by the lack of a specific model for how or why protein neutrality might fluctuate.

More recently, researchers have preferred to describe neutral evolution using the concept of "neutral networks," which are networks in the space of possible protein sequences in which each functional protein is linked to all other functional proteins that differ by only a single mutation (SMITH 1970; HUYNEN *et al.* 1996; GOVINDARAJAN and GOLDSTEIN 1997; VAN NIMWEGEN *et al.* 1999; BORNBERG-BAUER and CHAN 1999; TIANA *et al.* 2000; BASTOLLA *et al.* 2002). A neutrally evolving protein population is then envisioned as moving on the neutral network, and the neutrality of the population may fluctuate if the nodes on the network differ in their connectivities. A general theoretical treatment of evolution on neutral networks by VAN NIMWEGEN *et al.* (1999) has shown that if the product of the population size and mutation rate is small then members of the population are equally likely to occupy any node, while if this product is large then the population will preferentially occupy highly connected nodes (see also (BORNBERG-BAUER and CHAN 1999; TAVERNA and GOLDSTEIN 2002b; XIA and LEVITT 2004)). Simulations with simplified lattice models of proteins have attempted to provide insight into the specific features of protein neutral networks. These simulations have shown that lattice protein neutral networks are centered around highly connected nodes occupied by stable proteins (BORNBERG-BAUER and CHAN 1999; XIA and LEVITT 2004; WINGREEN *et al.* 2004; BROGLIA *et al.* 1999), a finding consistent with the experimental observation (BLOOM *et al.* 2005;

BLOOM *et al.* 2006) that stable proteins are more mutationally robust. Lattice protein studies also suggest that protein structures differ in their "designabilities" (defined as the total number of sequences that fold into a structure), and that sequences that fold into more designable structures will neutrally evolve at a faster rate due to the increased size and connectivity of their neutral networks (GOVINDARAJAN and GOLDSTEIN 1997; LI *et al.* 1996; ENGLAND and SHAKHNOVICH 2003; CHAN and BORNBERG-BAUER 2002; WINGREEN *et al.* 2004). Finally, simulations have demonstrated that fluctuations in neutrality as a protein population moves along its neutral network can lead to an overdispersion of the molecular clock (BASTOLLA *et al.* 2002), as originally suggested by Takahata. However, an extension of these lattice protein simulations of evolution on neutral networks into a quantitative theory has been difficult because protein neutral networks are far too large to be computed for all but the simplest lattice models.

Here we present a mathematical treatment of neutral protein evolution that describes the evolutionary dynamics in terms of the $\Delta\Delta G$ values for single mutations, which are experimentally measurable. Our treatment is based on the experimentally verified (BLOOM *et al.* 2005; BLOOM *et al.* 2006) connection between protein stability and mutational robustness, as well as a few biophysically supported assumptions about $\Delta\Delta G$ values for random mutations. By linking a protein's tolerance to mutations with stability, we are able to quantitatively describe neutral evolution without a full description of the neutral network. We can then compute the average number of accumulated mutations, the average fraction of neutral mutations, the index of dispersion, and the distribution of stabilities in a neutrally evolving population solely from knowledge of the $\Delta\Delta G$ values for single mutations. In addition, we follow the formalism of

VAN NIMWEGEN *et al.* (1999) to calculate all four of these properties in the limit when the product of the population size and mutation rate is much less than one and in the limit when this product is much greater than one. In demonstrating that these properties are different in these two limits, we show that the rate of fixation of neutral mutations can vary with population size in violation of one of the standard predictions of Kimura's neutral theory (KIMURA 1987). Our work presents a unified view of neutral protein evolution that is grounded in measureable thermodynamic quantities.

MATERIALS AND METHODS

Lattice Protein Simulations We performed simulations with lattice proteins of $L = 20$ monomers of 20 types corresponding to the natural amino acids. The proteins could occupy any of the 41,889,578 possible compact or non-compact conformations on a two-dimensional lattice. The energy of a conformation $\mathcal{C}$ is the sum of the nonbonded nearest-neighbor interactions, $E\left(\mathcal{C}\right) = \sum_{i=1}^{L} \sum_{j=1}^{i-2} C_{ij}\left(\mathcal{C}\right) \times \epsilon\left(\mathcal{A}_i, \mathcal{A}_j\right)$, where $C_{ij}\left(\mathcal{C}\right)$ is one if residues i and j are nearest neighbors in conformation $\mathcal{C}$ and zero otherwise, and $\epsilon\left(\mathcal{A}_i, \mathcal{A}_j\right)$ is the interaction energy between residue types $\mathcal{A}_i$ and $\mathcal{A}_j$, given by Table 5 of MIYAZAWA and JERNIGAN (1985). We computed the stability of a conformation $\mathcal{C}_t$ as $\Delta G_f\left(\mathcal{C}_t\right) = E\left(\mathcal{C}_t\right) + T \ln\left\{Q\left(T\right) - \exp\left[-E\left(\mathcal{C}_t\right)/T\right]\right\}$, where $Q\left(T\right) = \sum_{\{\mathcal{C}_i\}} \exp\left[-E\left(\mathcal{C}_i\right)/T\right]$ is the partition sum, made tractable by noting that there are only 910,972 unique contact sets. All simulations were performed at a reduced temperature of $T = 1.0$

We used adaptive walks to find sequences that folded into each of the three arbitrarily chosen conformations shown in Fig. 2 with $\Delta G_f \leq 0$, and then neutrally evolved these sequences

8

for 10^4 generations with a population size of $N = 100$. Our evolutionary algorithm was as follows: at each generation we randomly chose a protein that folded to the parental structure with $\Delta G_f \leq 0$ from the population and mutated each residue to some other randomly chosen residue with probability 5×10^{-4}, and continued doing this until we had filled the new population with proteins. At the end of this equilibration evolution, we chose the most abundant sequence in the population as the starting point for further analysis and for the computation of the distribution of $\Delta\Delta G$ values for all 380 point mutations (sequences shown in Fig. 2). In principle, computing the distribution of $\Delta\Delta G$ values over all sequences in the population rather than just the most abundant one should give a more accurate representation of the true form of this distribution, and indeed we found that doing this slightly increased the accuracy of the predictions shown in Fig. 2. However, the resulting improvement in accuracy was small, since the approximate constancy of the $\Delta\Delta G$ distribution during neutral evolution (discussed below) means that the distribution computed over a single sequence is representative of that computed over all sequences in the population. Therefore, we chose to compute the $\Delta\Delta G$ distribution over just the most abundant sequence since this choice more closely tracks what would be experimentally feasible with real proteins. (It is experimentally tractable to compute $\Delta\Delta G$ values for a single protein, but would be unmanageable to do so for all proteins in a natural population.)

To collect data for the case when the product $N\mu$ of the population size N and the per protein per generation mutation rate μ is $\ll 1$, we first equilibrated 1,000 replicates by evolving each of them with a population size of $N = 10$ and for 5,000 generations starting with

a clonal population of the initial sequence described above. The remainder of the evolution-ary algorithm was as described above: the mutation rate stayed at 5×10^{-4} per residue per generation (corresponding to a per protein per generation mutation rate of $\mu = 10^{-2}$), and at each generation all proteins that folded to the target native structure with $\Delta G_f \leq 0$ reproduced with equal probability. We then evolved each of these equilibrated populations for a further 5,000 generations to collect data. We combined the data for all the folded proteins in the fi-nal populations of all the replicates to calculate the average number of mutations $\langle m \rangle_T$ after T generations, the corresponding index of dispersion R_T, and the distribution of stabilities shown in Fig. 2. If we instead simply randomly chose a single folded protein from the final popu-lation of each replicate, we obtained results that were identical within the precision shown in Fig. 2. We emphasize that $\langle m \rangle_T$ and R_T were computed by keeping track of the actual number of mutations that had occurred during the evolutionary history of each protein, not simply by counting the number of amino acid differences between the ancestral and final sequences (the two quantities may differ if a single site undergoes multiple mutations, as discussed in more detail in later sections).

To generate the data for $N\mu \gg 1$, we used the same procedure but with $N = 10^5$ and only performed 10 replicates. We again computed the statistics shown in Fig. 2 by combining the data for all of the folded proteins in the final populations of all 10 replicates. Similar results were obtained if we instead computed $\langle m \rangle_T$ and R_T over all of the folded proteins in the final population of a single replicate (average values of $\langle m \rangle_T$ were identical while the R_T values of 1.03, 0.95, and 0.94 were extremely similar to those shown from top to bottom in Fig. 2). This

outcome is expected since the probability distributions for $N\mu \gg 1$ evolve deterministically.

Lattice Protein Predictions The numerical predictions for the lattice proteins given in Fig. 2 were computed by constructing the matrix $\mathbf{W}$ described in the first section of RESULTS with a bin size of $b = 0.005$ and truncating the matrix by assuming that no proteins would have stabilities less than -5.0. For the case when $N\mu \ll 1$, $\langle m \rangle_T$ was calculated using Equation 6 and R_T was calculated using Equation 11. For $N\mu \gg 1$, $\langle m \rangle_T$ was calculated using Equation 18 and R_T was calculated using Equation 19.

<div align="center">RESULTS</div>

Assumptions and Mathematical Background In this section we describe the physical view of protein evolution that motivates our work. We begin with the basic observations that evolution selects for protein function, and that most proteins must stably fold in order to function (ANFINSEN 1973), meaning that protein stability is under evolutionary pressure only insofar as it must be sufficient to allow a protein to fold and function. In taking this view, we ignore those proteins (estimated at 10% of prokaryotic and 30% of eukaryotic proteins) that are intrinsically disordered (UVERSKY *et al.* 2005), as well as those rare proteins that are only kinetically stable (JASWAL *et al.* 2002). Natural selection for function requires a protein to fold with some minimal stability $\Delta G_f^{\min}$, since proteins that lack this minimal stability will be unable to reliably adopt their native structure and perform their biochemical task. A protein's extra stability beyond this minimal threshold is quantified as $\Delta G_f^{\text{extra}} = \Delta G_f - \Delta G_f^{\min}$, meaning that all functional proteins must have $\Delta G_f^{\text{extra}} \leq 0$ (more negative val-

ues of ΔG_f indicate increased stability). We further assume that as long as $\Delta G_f^{\text{extra}} \leq 0$, natural selection for protein function is indifferent to the exact amount of extra stability a protein posesses. This assumption is at odds with the persistent speculation that high stability inherently impairs protein function and so is selected against by evolution (DEPRISTO *et al.* 2005; SOMERO 1995). But the circular argument most commonly advanced to support this speculation — that the observed marginal stability of natural proteins indicates that higher stability is detrimental to protein function — has now been contradicted both by experiments that have dramatically increased protein stability without sacrificing function (SERRANO *et al.* 1993; GIVER *et al.* 1998; VAN DEN BURG *et al.* 1998; ZHAO and ARNOLD 1999) and by demonstrations that marginal stability is a simple consequence of the fact that most mutations are destabilizing (TAVERNA and GOLDSTEIN 2002a; ARNOLD *et al.* 2001, as well as the current work). There is a possibility, however, that certain regulatory proteins must be marginally stable to faciliate rapid degradation (HUNTZICKER *et al.* 2006). To summarize, current biochemical evidence supports our assumption that (with certain well-defined exceptions) the only requirement imposed on protein stability by natural selection for protein function is that stability must meet or surpass some minimal threshold (a protein must have $\Delta G_f^{\text{extra}} \leq 0$).

A mutation to a protein changes its stability by an amount $\Delta\Delta G$, and experimental measurements of $\Delta\Delta G$ values have shown that most mutations are destabilizing (have $\Delta\Delta G > 0$) (PAKULA *et al.* 1986; GODOY-RUIZ *et al.* 2004; MATTHEWS 1993; KUMAR *et al.* 2006). A mutation is neutral with respect to selection for stability if $\Delta\Delta G + \Delta G_f^{\text{extra}} \leq 0$ since the mutant protein still satisfies the minimal stability threshold; otherwise the mutant does

not stably fold and is culled by natural selection. Of course, mutations can also have specific effects on protein function (such as altering an enzyme's activity), but experiments have shown that such mutations are rare compared to the large number of mutations that affect stability (SHORTLE and LIN 1985; PAKULA *et al.* 1986; LOEB *et al.* 1989; BLOOM *et al.* 2006). Mutations can also have effects unrelated to the functioning of the individual protein molecule: they can affect its propensity to aggregate (CHITI *et al.* 2000), alter its codon usage (AKASHI 2003), change its mRNA stability (CHAMARY and HURST 2005), affect the efficiency or accuracy of translation (AKASHI 2003; ROCHA and DANCHIN 2004), or change the fraction of mistranslated proteins that fold (DRUMMOND *et al.* 2005). These higher-level effects are probably most apparent in the evolution of highly expressed proteins (DRUMMOND *et al.* 2005; PAL *et al.* 2001). However, here we ignore such effects and assume that the evolutionary impact of a mutation is mostly determined by its effect on protein stability (an assumption in agreement with a recent bioinformatics analysis by SANCHEZ *et al.* (2006)). The view we present therefore describes the impact of a mutation solely by its $\Delta\Delta G$ value and the $\Delta G_f^{\text{extra}}$ of the wildtype protein, and is summarized graphically in Fig. 1 We have previously used a similar view to successfully describe experimental protein mutagenesis results (BLOOM *et al.* 2005; BLOOM *et al.* 2006).

[Figure 1 about here.]

To use the view of Fig. 1 to construct a useful description of neutral protein evolution, we make one major assumption: that the overall distribution of $\Delta\Delta G$ values for random mutations stays roughly constant as the protein sequence evolves. Actually, this assumption is

13

stronger than is strictly needed for the mathematical theory presented below — the theory can be developed simply by assuming that all proteins with the same ΔG_f have the same distribution of $\Delta\Delta G$ values (in this case the matrix elements W_{ij} introduced below depend on j in addition to the difference $i - j$). However, we make the stronger assumption that the $\Delta\Delta G$ distribution remains constant during sequence evolution, since we believe that this assumption is consistent with existing evidence. We emphasize that this assumption does not imply that we are arguing that the $\Delta\Delta G$ distribution is identical for every possible protein sequence. Clearly, for any given structure there is a most stable sequence (with all $\Delta\Delta G$ values positive), a least stable sequence (with all $\Delta\Delta G$ values negative), and a vast range of sequences in between. However, most of these sequences fall within a stability range that is never populated by evolution, since simulations (TAVERNA and GOLDSTEIN 2002a) and experiments (KEEFE and SZOSTAK 2001; DAVIDSON *et al.* 1995) clearly show that the vast majority of protein sequences do not stably fold into any structure (meaning the least stable folded protein is still far more stable than the typical random sequence). Among the subset of sequences that do stably fold, the simple statistical reality that marginally stable sequences are far more abundant than highly stable sequences causes evolution to further confine itself mostly to sequences with stabilities far less than that of the most stable sequence (TAVERNA and GOLDSTEIN 2002a; ARNOLD *et al.* 2001, as well as the current work). This fact is amply demonstrated by engineering experiments that have greatly increased the stability of natural proteins without sacrificing any of their functional properties (SERRANO *et al.* 1993; GIVER *et al.* 1998; VAN DEN BURG *et al.* 1998; ZHAO and ARNOLD 1999). Therefore, al-

though the distribution of $\Delta\Delta G$ values certainly varies widely among all sequences, it is still reasonable to assume that it is relatively constant among those sequences visited by natural evolution. This assumption of a constant $\Delta\Delta G$ distribution among evolved sequences is explicitly supported by simulations (BLOOM *et al.* 2005; BLOOM *et al.* 2006; BROGLIA *et al.* 1999; WILKE *et al.* 2005, as well as the current work), and is consistent with the observation that the number of neighbors on a protein's neutral network is approximately determined by its stability (BORNBERG-BAUER and CHAN 1999; XIA and LEVITT 2004). Furthermore, protein mutagenesis experiments indicate that the $\Delta\Delta G$ values for random mutations are usually additive (SERRANO *et al.* 1993; WELLS 1990), meaning that any given mutation to a protein of length L will alter only $\approx 1/L$ of the other $\Delta\Delta G$ values, leaving the $\Delta\Delta G$ distribution mostly unchanged. Finally, the assumption of a constant $\Delta\Delta G$ distribution has been shown to explain the experimentally observed exponential decline in the fraction of functional proteins with increasing numbers of mutations (BLOOM *et al.* 2005). However, we acknowledge that at present the assumption of a roughly constant $\Delta\Delta G$ distribution among neutrally evolving proteins can be verified only for lattice proteins — for real proteins the most we can say is that it is consistent with existing experimental evidence.

We begin our mathematical treatment by conceptually dividing the continuous variable of protein stability into small discrete bins of width b. This discretization of stability allows us to treat mutations as moving a protein from one bin to another — the bins can be made arbitrarily small to eliminate any numerical effects of the binning. The stability of each folded protein in the evolving population (the folded proteins are all those with $\Delta G_f^{\text{extra}} \leq 0$) can be described

by specifying its stability bin. Specifically, a protein is in bin i if it has $\Delta G_f^{\text{extra}}$ between $(1-i)\,b$ and $-ib$, where $i = 1, 2, \ldots$. Let W_{ij} be the probability that a random mutation has a $\Delta\Delta G$ value such that it moves a protein's stability from bin j to bin i, where i and j both are in the range $1, 2, \ldots$. Then W_{ij} is easily computed as the fraction of $\Delta\Delta G$ values between $b\,(j - i - 1)$ and $b\,(j - i)$. Since W_{ij} only describes transitions between folded proteins, and since we have assumed that a protein's mutational tolerance is determined by its stability, then the fraction of folded mutants (neutrality) of a protein in bin j is $\nu_j = \sum_i W_{ij}$. Clearly, more stable proteins will have larger values of ν_j.

In the next two sections, we will use the matrix $\mathbf{W}$ with elements W_{ij} to calculate the distribution of stabilities in an evolving protein population of constant size N, the mean number of mutations $\langle m \rangle_T$ after T generations, the corresponding index of dispersion $R_T = \frac{\langle m^2 \rangle_T - \langle m \rangle_T^2}{\langle m \rangle_T}$, and the average fraction of mutations $\langle \nu \rangle$ that do not destabilize the proteins past the minimal stability threshold. We assume that $\mathbf{W}$ is computed from the distribution of $\Delta\Delta G$ values for all random single amino-acid mutations, although in principle it could be for any type of mutation. We also assume that the per-protein-per-generation mutation rate μ is small, so that at each generation a protein undergoes at most one mutation. Our calculations at first follow, and then extend the theoretical treatment by VAN NIMWEGEN *et al.* (1999) of evolution on a neutral network. In particular, we follow their lead in separately treating the two limiting cases where the product $N\mu$ of the population size and mutation rate is $\ll 1$ and $\gg 1$. We emphasize that all of the equations derived in the next two sections depend only on the mutation rate μ, the number of generations T, and the matrix $\mathbf{W}$ which can be computed

from the single-mutant $\Delta\Delta G$ values. The population size N determines the applicable limiting case, but otherwise drops out of all final results.

Limit when $N\mu \ll 1$ When $N\mu \ll 1$, the evolving population is usually clonal, since each mutation is either lost or goes to fixation before the next mutation occurs. If a mutation destabilizes a protein in the population beyond the stability cutoff, then it is immediately culled by natural selection. If a mutation does not destabilize a protein beyond the stability cutoff, it will be lost to genetic drift with probability $\frac{N-1}{N}$ and go to fixation with probability $1/N$ (KIMURA 1983). Since mutations occur rarely ($N\mu \ll 1$), the loss or fixation of the mutant will occur before the next mutant appears in the population. The entire population therefore moves as one entity along its neutral network. The population can thus be described by the column vector $\mathbf{p}(t)$, with element $p_i(t)$ giving the probability that the population is in stability bin i at time t.

If the population is initially in stability bin j, at each generation there is a probability $N\mu W_{ij}$ that a protein experiences a mutation that changes its stability to bin i, and if such a mutation occurs, then there is a probability of $1/N$ that it is eventually fixed in the population. Therefore, at each generation there is a probability μW_{ij} that the population experiences a mutation that eventually causes it to move from stability bin j to bin i. If we define the matrix $\mathbf{V}$ so that the diagonal elements are given by $V_{ii} = \nu_i$ and all other elements are zero, then $\mathbf{p}$ evolves according to

$$\mathbf{p}(t+1) \;\; = \;\; (\mathbf{I} - \mu\mathbf{V} + \mu\mathbf{W})\,\mathbf{p}(t) \tag{1}$$

where $\mathbf{I}$ is the identity matrix. Note that this equation treats lethal mutations (those that desta-bilize a protein beyond the cutoff) as immediately being lost to natural selection and so leaving the population in its original stability bin (hence the population accumulates a mutation with probability $\mu\mathbf{V}$ rather than probability μ). Equation 1 describes a Markov process with the non-negative, irreducible, and acyclic transition matrix $\mathbf{A} = \mathbf{I} - \mu\mathbf{V} + \mu\mathbf{W}$, and so $\mathbf{p}$ approaches the unique stationary distribution $\mathbf{p_o}$ satisfying

$$0 = (\mathbf{V} - \mathbf{W})\,\mathbf{p_o}. \tag{2}$$

This equation gives the expected distribution of protein stabilities solely in terms of the single-mutant $\Delta\Delta G$ values.

We now calculate the average number of mutations $\langle m \rangle_{T,o}$ that accumulate in an equili-brated population after T generations and the corresponding index of dispersion $R_{T,o}$. We em-phasize that $\langle m \rangle_{T,o}$ represents the average number of accumulated mutations during the course of the evolutionary process. When the number of accumulated mutations m is small compared to the length of the protein sequence L ($m \ll L$), then m is just equal to the number of residues differing from those in the parent protein sequence (the Hamming distance). However, when m becomes substantial relative to L, m becomes larger than the Hamming distance since some sites will undergo multiple mutations (JUKES and CANTOR 1969). In this case it is necessary to use a substitution model to infer m from the observed Hamming distance. In the treatment that follows, we calculate the expected value of m; application of these formulae to actual pro-tein sequences requires use of one of the well-established statistical techniques for inferring m from the Hamming distance (JUKES and CANTOR 1969; GOLDMAN and YANG 1994). We

begin the calculation of $\langle m \rangle_{T,o}$ by defining $\mathbf{p}(m,t)$ to be the column vector with element i giving the probability that at time t the population has accumulated m mutations and is in stability bin i. The time evolution of $\mathbf{p}(m,t)$ is given by

$$\mathbf{p}(m,t+1) = (\mathbf{I} - \mu\mathbf{V})\mathbf{p}(m,t) + \mu\mathbf{W}\mathbf{p}(m-1,t). \tag{3}$$

The kth moment of the number of mutations at time t is

$$\langle m^k \rangle_t = \mathbf{e}\sum_m m^k\mathbf{p}(m,t), \tag{4}$$

where $\mathbf{e} = (1,\ldots,1)$ is the unit row vector. We can write a recursive equation for $\langle m \rangle_t$ in the long-time limit (steady state) by multiplying both sides of Equation 3 by m, summing over m, and left multiplying by $\mathbf{e}$ to obtain

$$
\begin{aligned}
\langle m \rangle_{t+1} &= \mathbf{e}(\mathbf{I}-\mu\mathbf{V})\sum_m m\mathbf{p}(m,t) + \mu\mathbf{e}\mathbf{W}\sum_m m\mathbf{p}(m-1,t) \\
&= \mathbf{e}\mathbf{A}\sum_m m\mathbf{p}(m,t) + \mu\mathbf{e}\mathbf{W}\mathbf{p_o} \\
&= \langle m \rangle_t + \mu\langle \nu \rangle_o,
\end{aligned}
\tag{5}
$$

where we have used the property $\mathbf{e}\mathbf{A} = \mathbf{e}$, noted that in the long-time limit $\sum_m \mathbf{p}(m,t) = \mathbf{p_o}$ and $\sum_m m\mathbf{p}(m-1,t) = \sum_m [(m-1)\mathbf{p}(m-1,t) + \mathbf{p}(m-1,t)] = \sum_m m\mathbf{p}(m,t) + \mathbf{p_o}$, and defined the average neutrality as $\langle \nu \rangle_o = \mathbf{e}\mathbf{W}\mathbf{p_o} = \mathbf{e}\mathbf{V}\mathbf{p_o}$. Summing the recursion yields the steady-state value for the number of accumulated mutations,

$$\langle m \rangle_{T,o} = T\mu\langle \nu \rangle_o. \tag{6}$$

To calculate the index of dispersion $R_{T,o} = \frac{\langle m^2 \rangle_{T,o} - \langle m \rangle_{T,o}^2}{\langle m \rangle_{T,o}}$, we need to find the second moment $\langle m^2 \rangle_{T,o}$. In a fashion analogous to the construction of Equation 5, we can write a

recursive expression for the long-time limit of $\langle m^2 \rangle_{T,o}$ as

$$
\begin{aligned}
\langle m^2 \rangle_{t+1} &= \mathbf{e}\,(\mathbf{I} - \mu\mathbf{V}) \sum_m m^2 \mathbf{p}\,(m,t) + \mu\mathbf{eW} \sum_m m^2 \mathbf{p}\,(m-1,t) \\
&= \mathbf{eA} \sum_m m^2 \mathbf{p}\,(m,t) + 2\mu\mathbf{eW} \sum_m m\mathbf{p}\,(m,t) + \mu\mathbf{eWp_o} \\
&= \langle m^2 \rangle_t + 2\mu\mathbf{eW} \left[\mathbf{A} \sum_m m\mathbf{p}\,(m,t-1) + \mu\mathbf{Wp_o} \right] + \mu\langle \nu \rangle_o \\
&= \langle m^2 \rangle_t + 2\mu^2 \mathbf{eW} \sum_{\tau=0}^{t-1} \mathbf{A}^\tau \mathbf{Wp_o} + \mu\langle \nu \rangle_o \qquad\qquad (7)
\end{aligned}
$$

where we have used the property (implicit in Equation 5) that in the long-time limit, $\sum_m m\mathbf{p}\,(m,t) = \mathbf{A} \sum_m m\mathbf{p}\,(m,t-1) + \mu\mathbf{Wp_o}$. Summing the recursion yields the following value for the long-time limit,

$$
\begin{aligned}
\langle m^2 \rangle_{T,o} &= T\mu\langle \nu \rangle_o + 2\mu^2 \mathbf{eW} \sum_{t=0}^{T-1} \sum_{\tau=0}^{t-1} \mathbf{A}^\tau \mathbf{Wp_o} \\
&= T\mu\langle \nu \rangle_o + 2\mu^2 \mathbf{eW} \sum_{t=1}^{T} (T-t)\,\mathbf{A}^{t-1} \mathbf{Wp_o} \\
&= T\mu\langle \nu \rangle_o + T\,(T-1)\,\mu^2 \langle \nu \rangle_o^2 + 2\mu^2 \mathbf{eW} \sum_{t=1}^{T} (T-t)\left(\mathbf{A}^{t-1} - \mathbf{Q}\right) \mathbf{Wp_o}, \quad (8)
\end{aligned}
$$

where we have made the substitution $\mathbf{eWQWp_o} = \langle \nu \rangle_o^2$ and noted that $\lim_{t\to\infty} \mathbf{A}^t = \mathbf{Q} = (\mathbf{p_o}, \ldots, \mathbf{p_o})$ since $\mathbf{A}$ is an irreducible, aperiodic, stochastic matrix (EWENS and GRANT 2005). This yields a value for the index of dispersion in the long-time limit of

$$
R_{T,o} = 1 - \mu\langle \nu \rangle_o + \frac{2\mu}{\langle \nu \rangle_o} \mathbf{eW} \sum_{t=1}^{T} \left(1 - \frac{t}{T}\right) \left(\mathbf{A}^{t-1} - \mathbf{Q}\right) \mathbf{Wp_o}. \qquad (9)
$$

The above equation is consistent with the generic equation for the index of dispersion given by CUTLER (2000a) and CUTLER (2000b), where we now give concrete expressions for the

20

variables ρ and $h(t)$ in Cutler's formula in terms of measureable quantitites, namely $\rho = \mu\langle\nu\rangle_o$ and $h(t) = \frac{\mu}{\langle\nu\rangle_o}\mathbf{eWA}^{t-1}\mathbf{Wp_o}$.

We can further simplify Equation 9 by performing spectral decompositions of $\mathbf{A}$ and $\mathbf{Q}$. Let $\lambda_1, \ldots, \lambda_K$ be the eigenvalues of $\mathbf{V} - \mathbf{W}$, and let $\mathbf{r}_1, \ldots, \mathbf{r}_K$ and $\mathbf{l}_1, \ldots, \mathbf{l}_K$ be the corresponding right and left eigenvectors, normalized so that $\mathbf{l}_i\mathbf{r}_j = 1$ if $i = j$ and 0 otherwise. These eigenvectors are also eigenvectors of the irreducible, acyclic, stochastic $\mathbf{A}$, and the corresponding eigenvalues are $1 - \mu\lambda_1, \ldots, 1 - \mu\lambda_K$, with Perron-Frobenius theorems guaranteeing that one eigenvalue (chosen here to be $1 - \mu\lambda_1$) is equal to one and all other eigenvalues have absolute values less than one. Then $\mathbf{r}_1$ and $\mathbf{l}_1$ are right and left eigenvectors of $\mathbf{Q}$ with eigenvalue 1 (i.e. $\mathbf{r}_1 = \mathbf{p_o}$ and $\mathbf{l}_1 = \mathbf{e}$), and all other eigenvalues of $\mathbf{Q}$ are zero. The spectral decompositions are therefore $\mathbf{Q} = \mathbf{r}_1\mathbf{l}_1$ and $\mathbf{A} = \mathbf{r}_1\mathbf{l}_1 + \sum_{i=2}^{K}(1 - \mu\lambda_i)\mathbf{r}_i\mathbf{l}_i$. Inserting these spectral decompositions into Equation 9, we find for the index of dispersion a value of

$$R_{T,o} = 1 - \mu\langle\nu\rangle_o + \frac{2\mu}{\langle\nu\rangle_o}\mathbf{eW}\sum_{t=1}^{T}\left(1 - \frac{t}{T}\right)\sum_{i=2}^{K}(1 - \mu\lambda_i)^{t-1}\mathbf{r}_i\mathbf{l}_i\mathbf{Wp_o}, \qquad (10)$$

since $\mathbf{A}^t = \mathbf{r}_1\mathbf{l}_1 + \sum_{i=2}^{K}(1 - \mu\lambda_i)^t\mathbf{r}_i\mathbf{l}_i$ (EWENS and GRANT 2005). In the limit of large T and small μ, the value of $R_{T,o}$ given by the above equation approaches the value

$$\begin{aligned}
R_{T,o} &\approx 1 + \frac{2\mu}{\langle\nu\rangle_o}\mathbf{eW}\sum_{t=1}^{T}\sum_{i=2}^{K}(1 - \mu\lambda_i)^{t-1}\mathbf{r}_i\mathbf{l}_i\mathbf{Wp_o} \\
&\approx 1 + \frac{2}{\langle\nu\rangle_o}\mathbf{eW}\sum_{i=2}^{K}\lambda_i^{-1}\mathbf{r}_i\mathbf{l}_i\mathbf{Wp_o}, \qquad (11)
\end{aligned}$$

where the $\mu\langle\nu\rangle_o$ term drops out because μ is small and the $\sum_{t=1}^{T}\frac{t}{T}(1 - \mu\lambda_i)^{t-1}$ term drops out because T is large and $|1 - \mu\lambda_i| < 1$. This equation shows that $R_{T,o}$ approaches a constant

value independent of T and μ. Although we could not prove that value of $R_{T,o}$ given by Equation 11 is necessarily greater than one (since some of the eigenvalues λ_i could be complex), in all of our simulations we observed $R_{T,o} > 1$, suggesting that when $N\mu \ll 1$, fluctuations in protein stability tend to overdisperse the molecular clock.

Limit when $N\mu \gg 1$ When $N\mu \gg 1$, the population is spread across many nodes of the neutral network rather than converged on a single sequence (VAN NIMWEGEN *et al.* 1999). In this limit, we treat the evolutionary dynamics of the population deterministically (i.e., we assume an infinite population size), and describe the distribution of stabilities in the population by the column vector $\mathbf{x}(t)$, with element $x_i(t)$ giving the fraction of proteins in the population at time t that have stabilities in bin i. At generation t, the fraction of mutated proteins that continue to fold is $\langle \nu \rangle_t = \mathbf{e}\mathbf{W}\mathbf{x}(t)$. These folded proteins reproduce, and in order to maintain a constant population size, this reproduction must balance the removal of proteins by death, meaning that each folded sequence must produce an average of $\alpha_t = [1 - \mu(1 - \langle \nu \rangle_t)]^{-1}$ offspring. The population therefore evolves according to

$$\mathbf{x}(t+1) = \alpha_t \left[(1-\mu)\mathbf{I} + \mu\mathbf{W}\right]\mathbf{x}(t). \tag{12}$$

After the population has evolved for a sufficient period of time, $\mathbf{x}$ approaches an equilibrium distribution of $\mathbf{x}_\infty$. The corresponding equilibrium neutrality is $\langle \nu \rangle_\infty = \mathbf{e}\mathbf{W}\mathbf{x}_\infty$, and the equilibrium reproduction rate is $\alpha = [1 - \mu(1 - \langle \nu \rangle_\infty)]^{-1}$, so

$$\mathbf{x}_\infty = \alpha \left[(1-\mu)\mathbf{I} + \mu\mathbf{W}\right]\mathbf{x}_\infty. \tag{13}$$

22

This equation can be rewritten to show that $\mathbf{x}_\infty$ is the principal eigenvector of $\mathbf{W}$,

$$\langle \nu \rangle_\infty \mathbf{x}_\infty = \mathbf{W}\mathbf{x}_\infty. \tag{14}$$

We note that $\langle \nu \rangle_\infty$ approximates the asymptotic neutrality for the decline in the fraction of folded proteins upon random mutagenesis (BLOOM *et al.* 2005; WILKE *et al.* 2005).

We now determine the average number of accumulated mutations $\langle m \rangle_{T,\infty}$ and the corresponding index of dispersion $R_{T,\infty}$ by treating the forward evolutionary process. As described in the text immediately prior to Equation 3, our calculations describe the actual number of mutations accumulated during the evolutionary process, which may differ from the number of sequence differences relative to the ancestor if a single site undergoes multiple mutations. When $N\mu \gg 1$, it is not *a priori* obvious that the average number of mutations present in the population is equivalent to number of fixed substitutions along the line of descent. Therefore, in the APPENDIX, we show that identical results are obtained by tracing a randomly chosen protein backwards in time along its ancestor distribution, proving the treatment we give below is mathematically equivalent to treating the time-reversed process. We define $\mathbf{x}(m, t)$ as the column vector with element i giving the fraction of the population at time t that has accumulated m mutations and is in stability bin i. Once the population has reached the equilibrium distribution of stabilities, the time evolution of $\mathbf{x}(m, t)$ is

$$\mathbf{x}(m, t+1) = \alpha(1 - \mu)\mathbf{x}(m, t) + \alpha\mu\mathbf{W}\mathbf{x}(m - 1, t). \tag{15}$$

The recursion can be solved to obtain

$$\mathbf{x}(m, t) = \alpha^t \sum_{\kappa=0}^{t} \binom{t}{\kappa}(1 - \mu)^{t-\kappa}\mu^\kappa\mathbf{W}^\kappa\mathbf{x}(m - \kappa, 0), \tag{16}$$

23

as can be verified by direct substitution. Since we are assuming the population has equilibrated

at time 0 and no mutations have accumulated at that time, $\mathbf{x}(m, 0)$ is $\mathbf{x}_\infty$ for $m = 0$ and 0

otherwise. Furthermore, $\mathbf{x}_\infty$ satisfies Equation 14, so multiplying Equation 16 by $\mathbf{e}$ yields

$$x(m, t) = \binom{t}{m} \alpha^t (1 - \mu)^{t-m} (\mu \langle \nu \rangle_\infty)^m, \tag{17}$$

where $x(m, t) = \mathbf{e}\mathbf{x}(m, t)$ gives the fraction of the population that has accumulated m muta-

tions after t generations. The average number of accumulated mutations after T generations is

the mean of this binomial distribution,

$$\langle m \rangle_{T,\infty} = \frac{T \mu \langle \nu \rangle_\infty}{1 - \mu (1 - \langle \nu \rangle_\infty)}. \tag{18}$$

Using the well known result for the variance of the binomial distribution, we find that the index

of dispersion is

$$R_{T,\infty} = 1 - \frac{\mu \langle \nu \rangle_\infty}{1 - \mu (1 - \langle \nu \rangle_\infty)}. \tag{19}$$

It is important to reiterate that the above equation was derived under the assumption that there

is at most one mutation per sequence per generation. For realistic distributions of mutations

(i.e. Poisson), this means that $\mu \ll 1$. In this regime, $R_{T,\infty}$ is close to one.

Lattice Protein Simulations We tested our theory's predictions on the evolutionary dynamics

of lattice proteins. Lattice proteins are simple protein models that are useful tools for study-

ing protein folding and evolution (CHAN and BORNBERG-BAUER 2002). Our lattice proteins

were chains of 20 amino acids that folded on a two-dimensional lattice. The energy of a lattice

protein conformation was equal to the sum of the pairwise interactions between non-bonded

amino acids (MIYAZAWA and JERNIGAN 1985). Each lattice protein has 41,889,578 possible conformations, and by summing over all of these conformations we could exactly determine the partition sum and calculate ΔG_f. We set a minimal stability threshold for the lattice proteins of $\Delta G_f^{\mathrm{min}} = 0$, meaning that we considered all proteins that folded to the target structure with $\Delta G_f \leq 0$ to be folded and functional, while all proteins with $\Delta G_f > 0$ were considered to be nonfunctional. We note that this stability threshold is equivalent to requiring a lattice protein to spend at least half of its time in the target native structure at equilibrium. We began by generating lattice proteins that stably folded to each of the three different structures shown in Figure 2. For each of these three proteins, we determined the distribution of $\Delta\Delta G$ values for all 380 single mutations (these distributions are shown in Figure 2). These distributions were used to construct the matrix $\mathbf{W}$ and to predict the equilibrium distribution of stabilities, the average number of mutations, and the indices of dispersion for both the $N\mu \ll 1$ and the $N\mu \gg 1$ cases, using the equations presented in the preceding sections.

To test the accuracy of these predictions, we then simulated evolving populations of the lattice proteins with a standard evolutionary algorithm using Wright-Fisher sampling. Briefly, the populations were held at a constant size of either $N = 10$ or $N = 10^5$. At each generation, a new population was created by choosing parents with equal probability from all folded proteins in the previous generation's population, and copying these parents into the new population with a mutation rate of 5×10^{-4} mutations per residue per generation. Since the proteins have a length of 20 amino acids, this mutation rate corresponds to a per-protein-per-generation mutation rate of $\mu = 10^{-2}$. Therefore, the product $N\mu$ is either 0.1 or 10^3, corresponding

to $N\mu \ll 1$ or $N\mu \gg 1$, respectively. We emphasize that the lattice protein evolutionary algorithm is the same for both population sizes. When $N = 10$ the population naturally follows dynamics approximating those presented for $N\mu \ll 1$, while when $N = 10^5$ it naturally follows dynamics approximating those presented for $N\mu \gg 1$ (as evidenced by the excellent agreement of the predictions with the simulations). For $N = 10$, we performed 1,000 replicates for each different structure. For $N = 10^5$, computational constraints limited us to 10 replicates for each structure (however the evolutionary dynamics are nearly deterministic in this case, so all replicates yielded similar results). We note that during the simulations we recorded the number of mutations that actually accumulated rather than simply computing the number of differences (Hamming distance) from the original sequence.

Figure 2 shows the theoretical predictions and simulation results for each of the three structures. The theoretical predictions are in good agreement with the simulation results. Figure 2 clearly shows that when $N\mu \gg 1$, the proteins tend to be more stable than when $N\mu \ll 1$. This extra stability is a biophysical manifestation of the neutrally evolved mutational robustness predicted by VAN NIMWEGEN *et al.* (1999). This increase in stability leads to a substantial increase number of accumulated mutations. In accordance with the theoretical predictions, when $N\mu \ll 1$ the index of dispersion is elevated above one by fluctuations in protein stability. Another clear results from the simulations is that proteins of different structure show markedly different distributions of stabilities and rates of sequence evolution due to the differences in their $\Delta\Delta G$ distributions. Overall, the simulations offer strong support for the validity of the theoretical predictions in the preceding sections.

[Figure 2 about here.]

DISCUSSION

We have presented that a theory that offers quantitative predictions about the distribution of stabilities, the average number of fixed mutations, and the index of dispersion for an evolving protein population in terms of the $\Delta\Delta G$ values for individual mutations. We have demonstrated that these predictions are accurate for simple lattice proteins, and have used existing biophysical evidence to argue that the basic theoretical assumptions should also be accurate for real proteins. In this section, we give qualitative interpretations of the mathematical results and discuss their implications for our understanding of protein evolution.

One major result is to show that the effects of protein structure on the rate of sequence evolution can be quantitatively cast in terms of the $\Delta\Delta G$ values for single mutations. Numerous lattice protein simulations have shown that protein structure can dramatically affect the rate of sequence evolution, since structures that are more "designable" (encoded by more sequences) can evolve their sequences more rapidly (as can be seen in Fig. 2 of this work) (GOVINDARAJAN and GOLDSTEIN 1997; BORNBERG-BAUER and CHAN 1999; TIANA *et al.* 2000; XIA and LEVITT 2004; LI *et al.* 1996; CHAN and BORNBERG-BAUER 2002; WINGREEN *et al.* 2004). Unfortunately, these simulations typically measure structural designability by enumerating a large number of lattice protein sequences, meaning that their findings cannot be extended to real proteins for which such extensive enumeration is impossible. However, recent theoretical work by ENGLAND and SHAKHNOVICH (2003) has made progress in connecting designability to observable structural properties, and a bioinformatics analysis based on this theoretical

27

measure of designability indicates that structure indeed influences the evolutionary rate of real proteins (BLOOM *et al.* 2006b). Our work provides a way to quantitatively relate the structural influences on protein evolution to experimentally measureable $\Delta\Delta G$ values, opening the door to further connecting structural designability and sequence evolution to laboratory stability measurements. Although thousands of $\Delta\Delta G$ values have been measured experimentally (KUMAR *et al.* 2006), at present there are no large sets of measurements for truly random mutations to a single protein. When such sets of measurements become available, it should be possible to use them in conjunction with the theory that we have presented to predict the neutralities of real proteins with different structures.

A second important result is to show that protein evolutionary dynamics can depend on the product of population size and mutation rate, $N\mu$. When $N\mu \gg 1$, the evolving protein population is polymorphic in stability and subject to frequent mutations, so the more stable (and thus more mutationally tolerant) proteins produce more folded offspring. In contrast, when $N\mu \ll 1$, the population is usually monomorphic in stability and so all members of the population are equally likely to produce folded offspring. The general tendency for populations to neutrally evolve mutational robustness when $N\mu \gg 1$ has previously been treated mathematically by VAN NIMWEGEN *et al.* (1999), and a variety of lattice protein simulations have noted the tendency of evolving protein populations to preferentially occupy highly connected neutral network nodes (BORNBERG-BAUER and CHAN 1999; TAVERNA and GOLDSTEIN 2002b; XIA and LEVITT 2004). Our work shows that for proteins, in the limiting cases when $N\mu \ll 1$ or $\gg 1$, this process can be rigorously described by considering only protein stability, rather

than requiring a full analysis of the neutral network (provided, as we have argued is likely to be the case, that the assumption of a roughly constant $\Delta\Delta G$ distribution holds for real proteins as well as it holds for our lattice proteins). In addition, we prove that the number of accumulated mutations depends on whether $N\mu$ is $\ll 1$ or $\gg 1$. This finding is at odds with the standard prediction (KIMURA 1987) of Kimura's neutral theory that the rate of evolution is independent of population size. The reason for this discrepancy is that the standard neutral theory fails to account for the possibility that increasing the population size so that $N\mu \gg 1$ can systematically increase the fraction of mutations that are neutral.

A third important contribution of our theory is to use the distribution of $\Delta\Delta G$ values for single mutations to predict the distribution of protein stabilities in an evolving population. Several researchers have pointed out that evolved proteins will be marginally stable simply because most mutations are destabilizing (TAVERNA and GOLDSTEIN 2002a; ARNOLD *et al.* 2001); we have described this process quantitatively. In addition, we have shown how the neutral evolution of mutational robustness when $N\mu \gg 1$ will shift the proteins towards higher stabilities (as shown in Fig. 2), although this increase in stability is limited by the counterbalancing pressure of predominantly destabilizing mutations. The formulae we provide can in principle be combined with experimentally measured $\Delta\Delta G$ values to predict the expected range of stabilities for evolved proteins.

Our work also weds Takahata's concept that fluctuating neutral spaces might overdisperse the molecular clock (TAKAHATA 1987; CUTLER 2000b; BASTOLLA *et al.* 2002) to a concrete discription of how protein neutrality fluctuates during evolution. When $N\mu \ll 1$, fluctuations

29

in protein stability can cause an overdispersion in the number of accumulated substitutions that can be calculated from the single-mutant $\Delta\Delta G$ distribution. Furthermore, given our assumption of a roughly constant $\Delta\Delta G$ distribution, we show that the index of dispersion will approach a constant value that is independent of time or mutation rate, but will depend on whether $N\mu \ll 1$ or $\gg 1$. Previous simulations have indicated that overdispersion indeed depends on the population size (BASTOLLA *et al.* 2002; WILKE 2004) — we have explained this dependence by showing that stability-induced overdispersion does not occur when $N\mu \gg 1$ since the population's distribution of stabilities equilibrates as it spreads across many sequences. Mathematically, the difference in the cases $N\mu \gg 1$ and $N\mu \ll 1$ is that, assuming the $\Delta\Delta G$ distribution remains relatively constant, when the population size is sufficiently large, the distribution of protein stabilities no longer fluctuates in a manner that influences the probability of a substitution (Equation 3 contains $\mu\mathbf{V}$ in the first term on the right side, while Equation 15 does not).

In summary, we have presented a mathematical theory of how thermodynamics shape neutral protein evolution. A major strength of our theory is that it makes quantitative predictions using single-mutant $\Delta\Delta G$ values, which can be experimentally measured. Our work also suggests how neutral and adaptive protein evolution may be coupled through protein thermodynamics. Protein stability represents an important hidden dimension in the evolution of new protein function, since extra stability that is itself neutral can allow a protein to tolerate mutations that confer new or improved functions (BLOOM *et al.* 2006). Our theory describes the dynamics of protein stability during neutral evolution — adaptive protein evolution is super-

imposed on these stability dynamics, with proteins most likely to acquire beneficial mutations when they are most stable.

ACKNOWLEDGEMENTS

LITERATURE CITED

AKASHI, H., 2003 Translational selection and yeast proteome evolution. Genetics **164:** 1291–1303.

ANFINSEN, C. B., 1973 Principles that govern the folding of protein chains. Science **181:** 223–230.

ARNOLD, F. H., P. L. WINTRODE, K. MIYAZAKI, and A. GERSHENSON, 2001 How enzymes adapt: lessons from directed evolution. Trends Biochem. Sci. **26:** 100–107.

BASTOLLA, U., M. PORTO, H. E. ROMAN, and M. VENDRUSCOLO, 2002 Lack of self-averaging in neutral evolution of proteins. Phys. Rev. Lett. **89:** 208101.

BLOOM, J. D., S. T. LABTHAVIKUL, C. R. OTEY, and F. H. ARNOLD, 2006 Protein stability promotes evolvability. Proc. Natl. Acad. Sci. USA **103:** 5869–5874.

BLOOM, J. D., D. A. DRUMMOND, F. H. ARNOLD, and C. O. WILKE, 2006b Structural

determinants of the rate of protein evolution in yeast. Mol. Biol. Evol. **23:** 1751–1761.

BLOOM, J. D., J. J. SILBERG, C. O. WILKE, D. A. DRUMMOND, C. ADAMI, and F. H. ARNOLD, 2005 Thermodynamic prediction of protein neutrality. Proc. Natl. Acad. Sci. USA **102:** 606–611.

BORNBERG-BAUER, E. and H. S. CHAN, 1999 Modeling evolutionary landscapes: mutational stability, topology, and superfunnels in sequence space. Proc. Natl. Acad. Sci. USA **96:** 10689–10694.

BROGLIA, R. A., G. TIANA, H. E. ROMAN, E. VIGEZZI, and E. SHAKHNOVICH, 1999 Stability of designed proteins against mutations. Phys. Rev. Lett. **82:** 4727–4730.

CHAMARY, J. V. and L. D. HURST, 2005 Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. Genome Biology **6:** R75.

CHAN, H. S. and E. BORNBERG-BAUER, 2002 Perspectives on protein evolution from simple exact models. Applied Bioinformatics **1:** 121–144.

CHITI, F., N. TADDEI, M. BUCCIANTINI, P. WHITE, G. RAMPONI, and C. M. DOBSON, 2000 Mutational analysis of the propensity for amyloid formation by a globular protein. EMBO J. **19:** 1441–1449.

CUTLER, D. J., 2000a The index of dispersion of molecular evolution: slow fluctuations. J. Theor. Population Biol. **57:** 177–186.

CUTLER, D. J., 2000b Understanding the overdispersed molecular clock. Genetics **154:** 1403–1417.

DAVIDSON, A. R., K. J. LUMB, and R. T. SAUER, 1995 Cooperatively folded proteins in random sequence libraries. Nat. Struct. Biol. **2:** 856–864.

DEPRISTO, M. A., D. M. WEINREICH, and D. L. HARTL, 2005 Missense meanderings in sequence space: a biophysical view of protein evolution. Nat. Rev. Genetics **6:** 678–687.

DRUMMOND, D. A., J. D. BLOOM, C. ADAMI, C. O. WILKE, and F. H. ARNOLD, 2005 Why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. USA **102:** 14338–14343.

ENGLAND, J. L. and E. I. SHAKHNOVICH, 2003 Structural determinant of protein designability. Phys. Rev. Lett. **90:** 218101.

EWENS, W. J. and G. R. GRANT, 2005 *Statistical methods in bioinformatics* (2 ed.). New York, NY: Springer.

FERSHT, A. R., 1999 *Structure and Mechanism in Protein Science*. New York: W. H. Freeman and Company.

GIVER, L., A. GERSHENSON, P. O. FRESKGARD, and F. H. ARNOLD, 1998 Directed evolution of a thermostable esterase. Proc. Natl. Acad. Sci. USA **95:** 12809–12813.

GODOY-RUIZ, R., R. PEREZ-JIMENEZ, B. IBARRA-MOLERO, and J. M. SANCHEZ-RUIZ, 2004 Relation between protein stability, evolution and structure as probed by carboxylic acid mutations. J. Mol. Biol. **336:** 313–318.

GOLDMAN, N. and Z. YANG, 1994 A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11:** 725–736.

33

GOVINDARAJAN, S. and R. A. GOLDSTEIN, 1997 The foldability landscape of model proteins. Biopolymers **42:** 427–438.

GUO, H. H., J. CHOE, and L. A. LOEB, 2004 Protein tolerance to random amino acid change. Proc. Natl. Acad. Sci. USA **101:** 9205–9210.

HERMISSON, J., O. REDNER, H. WAGNER, and E. BAAKE, 2002 Mutation-selection balance: ancestry, load, and maximum principle. Theor. Pop. Biol. **62:** 9–46.

HUNTZICKER, E. G., I. S. ESTAY, H. ZHEN, L. A. LOKTEVA, P. K. JACKSON, and A. E. ORO, 2006 Dual degradation signals control Gli protein stability and tumor formation. Genes and Development **20:** 276–281.

HUYNEN, M. A., P. F. STADLER, and W. FONTANA, 1996 Smoothness within ruggedness: the role of neutrality in adaptation. Proc. Natl. Acad. Sci. USA **93:** 397–401.

JASWAL, S. S., J. L. SOHL, J. H. DAVIS, and D. A. AGARD, 2002 Energetic landscape of α-lytic protease optimizes longevity through kinetic stability. Nature **415:** 343–347.

JUKES, T. H. and C. R. CANTOR, 1969 *Mammalin protein metabolism, III*, Chapter Evolution of protein molecules, pp. 21–132. New York, NY: Academic Press.

KEEFE, A. D. and J. W. SZOSTAK, 2001 Functional proteins from a random-sequence library. Nature **410:** 715–718.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge, U.K.: Cambridge Univ. Press.

KIMURA, M., 1987 Molecular evolutionary clock and the neutral theory. J. Mol. Evol. **26:**

24–33.

KUMAR, M. D., K. A. BAVA, M. M. GROMIHA, P. PARABAKARAN, K. KITAJIMA, H. UEDAIRA, and A. SARAI, 2006 ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. Nucleic Acids Res **34:** D204–D206.

LESK, A. M. and C. CHOTHIA, 1980 How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. J. Mol. Biol. **136:** 225–270.

LI, H., R. HELLING, C. TANG, and N. WINGREEN, 1996 Emergence of preferred structures in a simple model of protein folding. Science **273:** 666–669.

LOEB, D. D., R. SWANSTROM, L. EVERITT, M. MANCHESTER, S. E. STAMPER, and C. A. H. III, 1989 Complete mutagenesis of the HIV-1 protease. Nature **340:** 397–400.

MATTHEWS, B. W., 1993 Structural and genetic analysis of protein stability. Annu. Rev. Biochem. **62:** 139–160.

MIYAZAWA, S. and R. L. JERNIGAN, 1985 Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules **18:** 534–552.

OHTA, T. and M. KIMURA, 1971 On the constancy of the evolutionary rate of cistrons. J. Mol. Evol. **1:** 18–25.

PAKULA, A. A., V. B. YOUNG, and R. T. SAUER, 1986 Bacteriophage λ *cro* mutations: effects on activity and intracellular degradation. Proc. Natl. Acad. Sci. USA **83:** 8829–

8833.

PAL, C., B. PAPP, and L. D. HURST, 2001 Highly expressed genes in yeast evolve slowly. Genetics **158:** 927–931.

ROCHA, E. P. C. and A. DANCHIN, 2004 An analysis of determinants of amino acids substitution rates in bacterial proteins. Mol. Biol. Evol. **21:** 108–116.

SANCHEZ, I. E., J. TEJERO, C. GOMEZ-MORENO, M. MEDINA, and L. SERRANO, 2006 Point mutations in protein globular domains: contributions from function, stability, and misfolding. J. Mol. Biol. **363:** 422–432.

SERRANO, L., A. G. DAY, and A. R. FERSHT, 1993 Step-wise mutation of barnase to binase: a procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. J. Mol. Biol. **233:** 305–312.

SHORTLE, D. and B. LIN, 1985 Genetic analysis of staphylococcal nuclease: identification of three intragenic "global" suppressors of nuclease-minus mutations. Genetics **110:** 539–555.

SMITH, J. M., 1970 Natural selection and the concept of a protein space. Nature **225:** 563–564.

SOMERO, G. N., 1995 Proteins and temperature. Annu. Rev. Physiol. **57:** 43–68.

TAKAHATA, N., 1987 On the overdispersed molecular clock. Genetics **116:** 169–179.

TAVERNA, D. M. and R. A. GOLDSTEIN, 2002a Why are proteins marginally stable? Proteins **46:** 105–109.

TAVERNA, D. M. and R. A. GOLDSTEIN, 2002b Why are proteins so robust to site mutations?

J. Mol. Biol. **315:** 479–484.

TIANA, G., R. A. BROGLIA, and E. I. SHAKHNOVICH, 2000 Hiking in the energy landscape in sequence space: a bumpy road to good folders. Proteins **39:** 244–251.

UVERSKY, V. N., C. J. OLDFIELD, and A. K. DUNKER, 2005 Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J. Mol. Recognit. **18:** 343–384.

VAN DEN BURG, B., G. VRIEND, O. R. VELTMAN, G. VENEMA, and V. G. H. EUSINK, 1998 Engineering an enzyme to resist boiling. Proc. Natl. Acad. Sci. USA **95:** 2056–2060.

VAN NIMWEGEN, E., J. P. CRUTCHFIELD, and M. HUYNEN, 1999 Neutral evolution of mutational robustness. Proc. Natl. Acad. Sci. USA **96:** 9716–9720.

WELLS, J. A., 1990 Additivity of mutational effects in proteins. Biochemistry **29:** 8509–8517.

WILKE, C. O., 2004 Molecular clock in neutral protein evolution. BMC Genetics **5:** 25.

WILKE, C. O., J. D. BLOOM, D. A. DRUMMOND, and A. RAVAL, 2005 Predicting the tolerance of proteins to random amino acid substitution. Biophysical J. **89:** 3714–3720.

WINGREEN, N. S., H. LI, and C. TANG, 2004 Designability and thermal stability of protein structures. Polymer **45:** 699–705.

XIA, Y. and M. LEVITT, 2004 Funnel–like organization in sequence space determines and distributions of protein stability and folding rate preferred by evolution. Proteins **55:** 107–114.

ZHAO, H. and F. H. ARNOLD, 1999 Directed evolution converts subtilisin E into a functional

equivalent of thermitase. Protein Eng. **12:** 47–53.

Zuckerkandl, E. and L. Pauling, 1965 *Evolving genes and proteins*, Chapter Evolutionary divergence and convergence in proteins, pp. 97–166. New York, NY: Academic Press.

APPENDIX

Here we calculate the properties of the evolving population when $N\mu \gg 1$ by analyzing the time-reversed process to compute the mean and variation in the number of mutations in a single randomly chosen protein over time. We show that the results so obtained are identical to those found in the main text, where we analyzed the forward-time process to compute the mean and variation in the number of mutations across the population of evolving proteins.

When $N\mu \gg 1$, the population is now never converged to a single sequence, so it is not *a priori* obvious that the average number of mutations present in the population is equivalent to the expected number of fixed substitutions along the line of descent. In fact, in the limit of very large population sizes there may not even be a common line of descent in relevant time frames, since many new mutations will occur before any given mutation goes to fixation. In the main text we calculated the average number of mutations $\langle m \rangle_{T,\infty}$ a sequence in the population has accumulated over the last T generations by treating the forward evolution of the population. Here we trace a randomly chosen protein in the population back in time, and show that the average number of substitutions $\langle s \rangle_T$ that it has accumulated over the last T generations is equal to $\langle m \rangle_{T,\infty}$. We also show that indices of dispersion of $\langle m \rangle_{T,\infty}$ and $\langle s \rangle_T$ have the same value of $R_{T,\infty}$.

To calculate $\langle s \rangle_T$, we first define a vector $\mathbf{a}$ giving the ancestor distribution (Hermisson *et al.* 2002):

element i of $\mathbf{a}\,(T-t)$ gives the probability that a randomly chosen sequence from the population at time T had a predecessor with stability in bin i at time $T-t$. The transition probabilities of $\mathbf{a}\,(T-t)$ when the population is in equilibrium are the discrete time analogue of those computed by HERMISSON *et al.* (2002). From Equation 15 of the main text, it follows that the fraction of sequences in bin i at time $t+1$ that had as their ancestor in the previous generation a sequence in bin j is $\alpha_t\,[(1-\mu)\,\delta_{ij}+\mu W_{ij}]\,x_j\,(t)$. In order to obtain the probability that a sequence in bin i at time $t+1$ had an ancestor in bin j, we have to divide this fraction by the total number of sequences in bin i at time $t+1$. When the population is at equilibrium, $\alpha_t=\alpha$ and $x_i\,(t+1)=x_i\,(t)=x_i$ where x_i is the element from $\mathbf{x}_\infty$. Hence, the probability that a sequence in bin i had an ancestor in bin j is $\alpha\,[(1-\mu)\,\delta_{ij}+\mu H_{ji}]$, where we have defined

$$H_{ji}=W_{ij}x_j/x_i, \tag{20}$$

The time evolution of $\mathbf{a}$ is therefore

$$\mathbf{a}\,(T-t)=\alpha\,[(1-\mu)\,\mathbf{I}+\mu\mathbf{H}]\,\mathbf{a}\,(T-t+1), \tag{21}$$

where the matrix $\mathbf{H}$ is defined by Equation 20. Equation 21 can be solved to show that the equilibrium value of $\mathbf{a}$ is $\mathbf{a}_\infty$ satisfying

$$\langle \nu \rangle_\infty \mathbf{a}_\infty = \mathbf{H}\mathbf{a}_\infty. \tag{22}$$

If we define $\mathbf{a}\,(s,T-t)$ as the vector with element i giving the probability that a randomly chosen sequence at time T had a predecessor at time $T-t$ in stability bin i and with s substitutions relative to the sequence at time T, then the time evolution for an equilibrated population

is

$$\mathbf{a}\left(s, T-t-1\right) = \alpha\left(1-\mu\right)\mathbf{a}\left(s, T-t\right) + \alpha\mu\mathbf{H}\mathbf{a}\left(s-1, T-t\right). \tag{23}$$

We can solve Equations 23 and 22 in a manner analogous to the forward process to obtain

$$\mathbf{a}\left(s, T-t\right) = \binom{t}{s}\alpha^t\left(1-\mu\right)^{t-s}\left(\mu\langle\nu\rangle_\infty\right)^s\mathbf{a}_\infty. \tag{24}$$

Again defining $a\left(s, T-t\right) = \mathbf{e}\mathbf{a}\left(s, T-t\right)$ as the probability of having accumulated s substitutions as one moves back t generations from time T, we obtain the binomial distribution

$$a\left(s, T-t\right) = \binom{t}{s}\alpha^t\left(1-\mu\right)^{t-s}\left(\mu\langle\nu\rangle_\infty\right)^s. \tag{25}$$

Comparison of Equation 17 of the main text and Equation 25 shows that they are identical. Therefore, all moments computed from the two distributions must be equal. In particular, this proves that $\langle m\rangle_{T,\infty} = \langle s\rangle_T$, and that the corresponding indices of dispersion have the same value of $R_{T,\infty}$ defined by Equation 19 of the main text. This shows that when $N\mu \gg 1$, we expect equivalent results regardless of whether we average over the number of mutations in all sequences present in the population, or randomly choose a single sequence and trace back along its ancestor distribution.
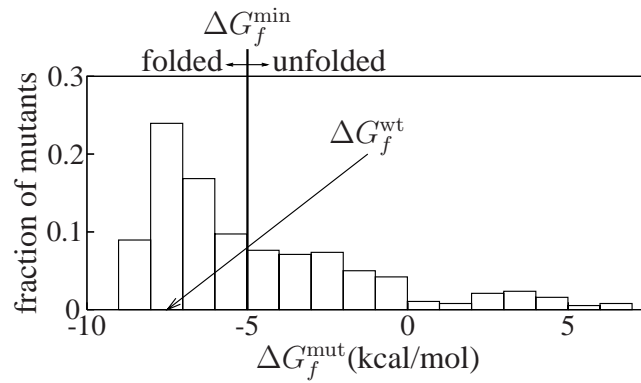
List of Figures

Figure 1: A thermodynamic view of protein evolution. A mutant protein stably folds if and only if it possesses some minimal stability, $\Delta G_f^{\min}$ (in this case -5 kcal/mol). The stability of the wildtype protein is $\Delta G_f^{\mathrm{wt}} = -7.5$ kcal/mol, meaning that it has $\Delta G_f^{\mathrm{extra}} = -2.5$ kcal/mol of extra stability. The bars show the distribution of $\Delta\Delta G$ values for mutations. Those mutants with $\Delta G_f^{\mathrm{extra}} + \Delta\Delta G \leq 0$ still stably fold, while all other mutants do not fold and so are culled by natural selection. The probability that a mutation will be neutral with respect to stable folding is simply the fraction of the distribution that lies to the left of the threshold. The data in this figure are hypothetical.
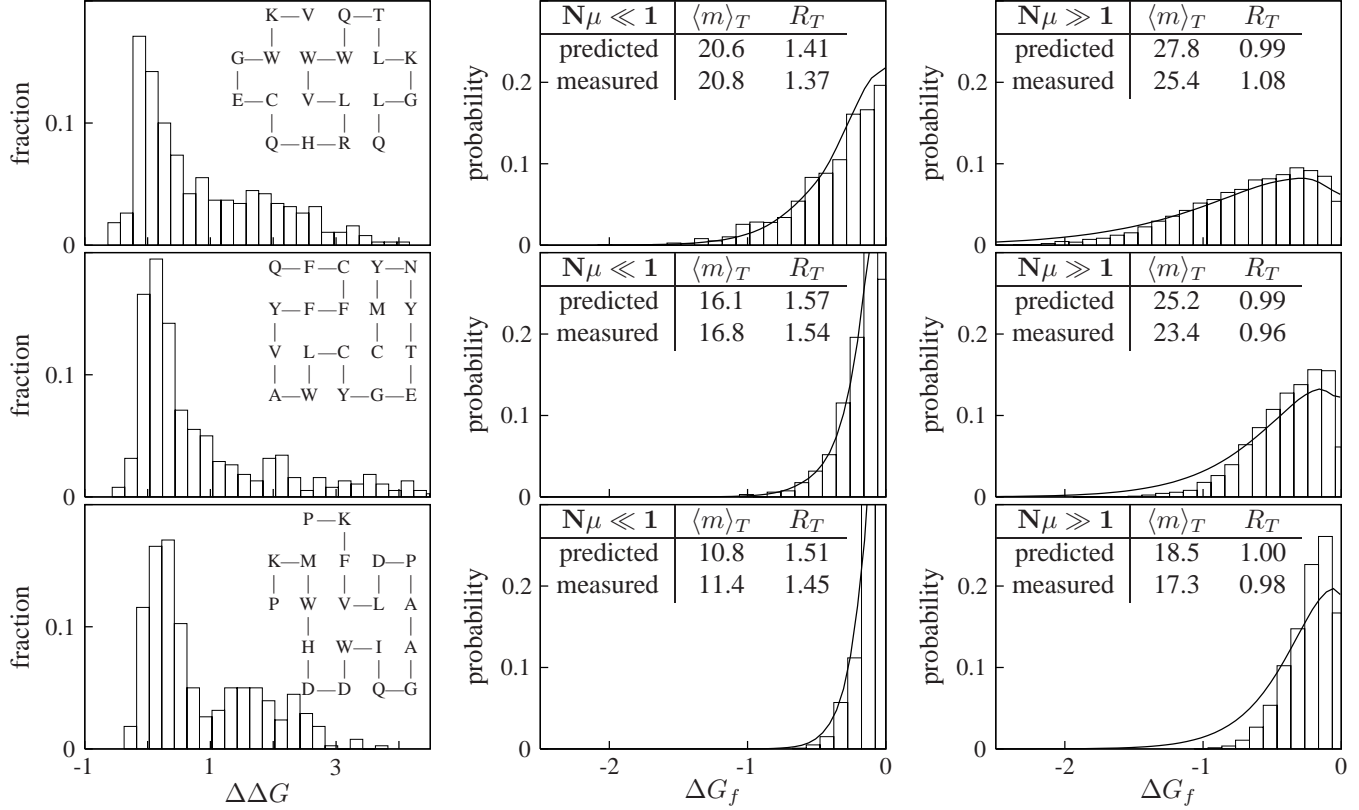
Figure 2: The theory gives accurate predictions for the evolution of model lattice proteins. Each row of panels corresponds to a different lattice protein. The graphs at left show the starting protein and the distribution of $\Delta\Delta G$ values for all point mutations. The graphs in the middle and right show the predicted (lines) and measured (boxes) distributions of stabilities among the evolved proteins. The tables embedded in the graphs show the predicted and measured values for the average number of mutations ($\langle m \rangle_T$) and the index of dispersion (R_T) after 5,000 generations of neutral evolution. The center graphs are for a population size of $N = 10$, and the graphs at the right are for $N = 10^5$. In both cases, the per protein per generation mutation rate is $\mu = 0.01$. As predicted, the evolving population with $N\mu \gg 1$ evolved mutational robustness that is manifested by increased protein stability. This additional mutational robustness accelerated the rate of sequence evolution.

Left panel axes: fraction, $\Delta\Delta G$ (-1, 1, 3); labels: K—H—R, E—H—A—L, F—E, H, Y—L, F—S, K—F, I, T, K—Q

Middle panel: probability, ΔG_f (-2, -1, 0)

$\mathbf{N}\mu \ll \mathbf{1}$	$\langle m \rangle_T$	R_T
predicted	9.7	1.49
measured	9.4	1.43

Right panel: probability, ΔG_f (-2, -1, 0)

$\mathbf{N}\mu \gg \mathbf{1}$	$\langle m \rangle_T$	R_T
predicted	16.1	1.00
measured	15.2	0.93